AD-A090 587 WEST VIRGINIA UNIV MORGANTOWN F/G 5/10
INFORMANT ACCURACY IN RECALL OF COMMUNICATIONS AND PERCEPTION O--ETC(U)
SEP 80 H R BERNARD, P D KILLWORTH, L SAILER N00014-75-C-0441
UNCLASSIFIED BK-122-80 NL

1 OF 1
AD
A090587

END
DATE
FILMED
11-80
DTIC

AD A090587

SECURITY CLASSIFICATION OF THIS PAGE (When Data Entered)

REPORT DOCUMENTATION PAGE		READ INSTRUCTIONS BEFORE COMPLETING FORM
1. REPORT NUMBER BK-122-80	2. GOVT ACCESSION NO. AD-A090 587	3. RECIPIENT'S CATALOG NUMBER
4. TITLE (and Subtitle) Informant accuracy in recall of communications and perception of one's place in global social structure and summary		5. TYPE OF REPORT & PERIOD COVERED Final Repts
		6. PERFORMING ORG. REPORT NUMBER
7. AUTHOR(s) H.R. Bernard, Peter D. Killworth, Lee Sailer		8. CONTRACT OR GRANT NUMBER(s) N00014-75-C-0441-P000 1
9. PERFORMING ORGANIZATION NAME AND ADDRESS H.R. Bernard, Dept. of Anthropology University of Florida, Gainesville, FL 32611		10. PROGRAM ELEMENT, PROJECT, TASK AREA & WORK UNIT NUMBERS
11. CONTROLLING OFFICE NAME AND ADDRESS ONR-Code 452 Arlington, VA 22217		12. REPORT DATE September 1, 1980
		13. NUMBER OF PAGES 21
14. MONITORING AGENCY NAME & ADDRESS(if different from Controlling Office) West Virginia University Morgantown, WV 26505		15. SECURITY CLASS. (of this report) Unlimited
		15a. DECLASSIFICATION/DOWNGRADING SCHEDULE

16. DISTRIBUTION STATEMENT (of this Report)

Approved for public release; distribution unlimited.

17. DISTRIBUTION STATEMENT (of the abstract entered in Block 20, if different from Report)

18. SUPPLEMENTARY NOTES

OCT 17 1980

A

19. KEY WORDS (Continue on reverse side if necessary and identify by block number)

Network analysis
Recall
Communications
Behavior & Cognition

20. ABSTRACT (Continue on reverse side if necessary and identify by block number)

This is a final update on the data tape for a major experiment on accuracy in human recall of communications. A full report of the experiment may be found in TR BK 119-79 and BK-120-80.

DD FORM 1 JAN 73 1473 EDITION OF 1 NOV 65 IS OBSOLETE
S/N 0102-014-6601

SECURITY CLASSIFICATION OF THIS PAGE (When Data Entered)

DDC FILE COPY

80 10 6

I. Introduction

This is the final report of contract #N000C014-75-C-0441-p00001, Code 452. We have conducted a series of experiments in two major areas of social network analysis. The first concerns the accuracy of data collected by asking people to recall their communications with others. Since virtually all theories of organizational structure, and of information diffusion are based on such data, it seemed important to know whether the data are accurate.

The second area of research tries to address directly the problem of structure in the diffusion of information. How does one acquire data about such structure? If we know the social structure (how people are related to one another), then we should be able to predict the spread of information and to interpret diffusion data, without relying on possibly inaccurate information.

We will summarize our findings here, taking each of the major areas separately.

II. Informant accuracy in social network data

a. Ultimately, many things, from theories about social structure to major policy decisions about community development programs, depend on the quality of fundamental data about the information diffusion process. Large scale behavioral studies of information diffusion are difficult to do. Consequently, scholars have tried to use communication recall data to describe the network along which information is assumed to flow. Many studies, following the classic example of Coleman, Katz, and Menzel (1957) treat communication recall networks as isomorphic with the social structure of a given group. But what if merely <u>asking</u> people whom they talk to produces inaccurate results? What if people honestly try to tell us (by rank ordering, or scaling, or just randomly recalling), their communications, but simply can not handle cognitively the amount of data required in order to do so accurately?

We have phrased this problem as follows: data about communication networks are collected by using an instrument. The instrument is a query, usually some form of "who do (did) you talk to over x period, and for how long and how often?" As a chemist might use a thermometer to test the temperature of a liquid, this instrument is "inserted" into a respondent; it is extracted, and data are recorded. A chemist would insist that the error bounds of the thermometer be known (i.e.,"above 200°C, for each 50 degrees, add .001 degree due to changes in the physical structure of the instrument"). Similarly, we ask "what are the error bounds of the instrument 'who do you talk to?'"

In an attempt to test the error bounds, we conducted a series of seven experiments. In each of these, we asked a variety of questions of the genre "who do you...?" in a variety of ways. We have always found that asking people who they talk to, and how much, produces totally inaccurate results. Furthermore, standard socio-economic indicators do not account for the inaccuracy. We have concluded two things. First, <u>people do not know, with any acceptable accuracy, with whom they communicate; in other words, recall of communication links in a network is not a proxy for communication behavior. Second, data manipulations which depend on respondents' ability to rank or scale accurately whom they talk to, are useless if what one wants is a description of behavioral social or communications structure</u>. The meaning of these conclusions for diffusion studies is clear: finding out, incorrectly, who people talk to by asking them, and then using the information to impute diffusion structures and flows, can only yield incorrect results.[1]

In the next two sections, we will describe the seven data sets from our experiments, and summarize our results. In subsection d) we will present a summary description of some structural analyses we have performed in order to try to improve the accuracy of our data. Then, in subsection e), we will offer some suggestions for how we might proceed to study information diffusion behaviorally.

A

b) The Data

Deaf 1

The first experiment (Killworth and Bernard, 1976) was conducted in 1975. We asked some of the members of a naturally-occurring group (the deaf owners of teletype machines in the Washington, D.C. telephone call area) to "rank the members of the group in the order in which you communicate with them." After answering our questions, the group logged their TTY communications until they had data for each of at least 21 days, thus providing a behavioral comparison for the ranking data. Thirty-one out of 32 persons in the group did the ranking, and 25 provided TTY logs. Of these 25, four had no communication with any other person in our sample during the logging period, and some people spent as much as three months before they logged 21 days of TTY use.

Firty-two percent of the time an individual was able to rank his or her first communicant first, second, third, or fourth. Forty-eight percent of the time our respondents did even worse than that. There was no significant difference between respondents in terms of how accurate they were. No obvious factors (e.g., gender, length of time they had owned a TTY, amount of communication, etc.) produced significant trends in accuracy. In order to account for only 40% of a person's total communication, and with only 75% reliability, his or her first 17 rankings (out of 31) had to be included. Accounting for 70% with 95% reliability required 24 rankings. These last findings made the entire ranking procedure seem pointless.

Deaf 2

Since our first experiment dealt with a group of deaf teletype users, we returned to this population for replication. Sixty members of the deaf community in Washington were selected randomly from amongst the then 387 registered teletype users. They were each presented with a list of all 387 persons in the "local deaf TTY community" and asked to select the persons with whom they believed they might communicate in the next month. Eventually, 54 respondents provided data, and they communicated with 594 different people on their TTYs. Twenty-eight of the 54 ranked the persons they chose, by "amount of communication," and the other 26 scaled those chosen from 1 to 5, or from "very little" to "a lot" of communication. Several criteria were used for ranking and scaling. These were a) amount of communication (in lines); b) frequency of communication (i.e., number of contracts with an individual); and c) importance of communication (a purely subjective measure). All ranking informants used criterion a); most used criterion b) and c) also (86, 89%, respectively). Virtually all scaling informants used all three criteria.[2]

Following the collection of these data, all 60 persons were asked to log their TTYs for a month, noting who they called and who called them, and how many lines of TTY output they generated on each call. As a result of illnesses and vacations, 28 of the rankers finished the logging, as did 26 of the scalers. The concatenation of these logs (54 in all) enables two sets of behavioral data to be calculated: (1) the amount of communication (in lines) between any two individuals (at least one of whom is among the 54), and (2) the frequency of communication between two individuals. Since members of families were treated as separate individuals, a large total of 594 different names eventually occurred in the behavioral data.

At the end of the month, each person was visited again. They were asked (1) to select from the deck of all registered TTY users those with whom they had communicated; and (2) to rank or scale those chosen. In this phase of the experiment, most of the participants felt it was too cumbersome to rank or scale on three criteria (amount, frequency, and importance of communication). They were thus asked to make their judgments on the basis of "amount of communication," i.e., how many lines of TTY output were generated. As in our earlier experiment, a few persons used video display units rather than TTYs. They logged in minutes, and this was converted, as in Deaf 1, with the corrected value of two minutes to one line. These data, and those in the next three experiments, are described in detail in Bernard and Killworth, 1977.

Hams

Our next set of data comes from a group of amateur radio operators (called "hams") in West Virginia, western Pennsylvania, and eastern Ohio. The hams belong to the Monongalia Wireless Association (MWA), which owns and maintains WR8ABM, a two-meter, FM repeater station.

With the cooperation of the MWA, we monitored all conversations on WR8ABM, around the clock for 27 days. This was done by using a voice-operated relay between a receiver and a tape recorder. By law, hams identify themselves with their "call" (a combination of letters and numbers) every ten minutes. Thus, all communicants could be monitored, and the length of their conversations (in minutes) could be recorded.

At the end of the 27-day monitoring period, a list of 54 users was drawn up who accounted for all but a small fraction of the air time. Each person was mailed a sheet with all 54 "calls," and asked to scale them from 0 to 9. A total of 44 usable responses were obtained.

This experiment yielded three sets of data: the amount of time any two persons were in contact; the number of times any two persons were in contact; and the 0-9 scales by 44 persons over the list of 54 users of the repeater.

Office

These data are from a small social science research firm (with 45 employees). This group is composed of several research project teams, each having senior staff, lower level assistants, clerks, and typists.

Recall, or cognitive data were collected from 40 persons; behavioral data were collected from 44 persons. At time 1, an observer walked through the office on four nonconsecutive workdays, covering the same ground every 15 minutes for five hours each working day. He noted every dyadic contact, including those contained in n-tuple conversations. At time 2, seven weeks later, the same observational procedure was followed. This was "mildly obtrusive" data collection. That is, the observer's presence was obvious, but he did not interact with the subjects actively.

Between times 1 and 2, each participant was given the familiar deck of cards containing the names of the other participants. They arranged (i.e., ranked) the cards from "most" to "least" on how often they talked to others in the office during a "normal working day." The question of frequency, amount, and importance of contact was raised often by the participants (they are, after all, social science researchers), but this was deliberately left vague. They were told to make up their own minds. Because their judgments were explicitly based on a "normal working day" the behavioral data from time 1 and 2 were aggregated here. They do differ significantly, but whether this is due to day-to-day fluctuation (which we do not define!) or to a systematic time variation in the group can not be answered easily.

Tech

The tech data, from our fifth experiment, are from a graduate program in technology education at West Virginia University. The program contains faculty, graduate students, and secretaries in three locations: two converted houses at the bottom of a hill, and a suite of offices "on the hill" in the main education building at the university. There are 37 persons in the program; three of these are on full-time field assignment over 100 miles from the university.

For one week a team of observers walked through the office spaces of the tech program. They covered the same ground every half hour, and noted all occurrences of persons in verbal contact. Any two persons in contact were scored. N-tuples were scored by dyads. The same comments on obtrusiveness apply as for the office data.

After a week of observation, each of the 34 persons on the main campus was handed a deck of cards containing the names of all other members of the group, and asked to rank the deck from "most to least communication that week." The question was purposely left rather vague; amount, frequency, or importance of communication was not specified. When the participant finished, he or she handed the deck to the experimenter. The experimenter then laid out the cards in order on a table in front of the participant. The participant was then asked if he or she wanted to make any changes in the order to reflect a "typical week's communication," as opposed to "last week's communication."

This experiment yielded three sets of data: the frequency of dyadic contact; the guesses at last week's communication; and the guesses at a typical week's communication.

Frat

Our sixth data set is a time-series in a college fraternity. The data consist of affective relations (how much i says he likes j); recall of communications (how much i says he talked to j over a period of 5 days); actual communication (from behavioral sampling, how much i did talk to j over the 5-day period) for all dyads in a closed group of 58.

Affect was collected on a scale of 1 (least like) to 11 (most like), and cognition on a scale of 1 (don't talk with) to 5 (talk with a great deal). Behavior was measured by an observer passing through the fraternity every 15 minutes for 21 hours a day, over a period of 5 days, at the end of which the affective and cognitive data were collected. Thus behavioral data exists on a 15-minute time scale. This entire procedure was repeated three times, separated by about 6 weeks in each case. These data are described and analyzed in Killworth and Bernard, 1978a.

EIES

Our seventh and final experiment examines the possibility that the inaccuracy we have found is a function of the time period over which informants are asked to recall their behavior. All the previous data sets were based on people recalling their communications during one of three "windows": the previous five days; the previous month; and the forthcoming month. Any period of time, or window, can be characterized by two quantities, which we call "lag" and "width." Width is the amount of time over which informants are asked to recall their behavior. Lag is the amount of time that has elapsed since the beginning of the window. Thus, the five-day windows in some of our previous experiments have a width of five days, and a lag of five days.

The majority of questions asked by students of social networks have a lag equal to the width, and a range of a few days to the lifetime of the informant. It seems plausible that very recent time windows should tend to be more accurate than windows far in the past. "Who did you talk to one minute ago?" should yield more accurate data than "who did you talk to for a minute at this time last month?". Similar variations in accuracy could be caused by different widths: "who did you talk to during a period of a week, a month ago?". The question addressed in this experiment is "what is the combination of lag and width which yields the most accurage social network data?"

This question was addressed using a computer based conferencing system called EIES (Electronic Information Exchange System). The New Jersey Institute of Technology developed the system under grants from the National Science Foundation. A complete description of EIES, including its technology and design philosophy may be found in Hiltz and Turoff (1978). Briefly, EIES allows an individual to exchange messages with others on the system by leaving the message in a central computer for pick-up during the next time the "receiver" logs on.

Between December, 1978 and April, 1979, 57 paid volunteer EIES users participated in our experiment. An invitation to participate in the experiment was sent to over 150 EIES members. Depending on the rate of their EIES use, each informant took up to 37 interviews, each for a specific lag and width. The informant was given a window and was then asked to

list the people with whom he or she communicated during that window. Next, informants were given an opportunity to add or to delete names from the list, and were asked to estimate the number of messages and the number of lines sent to and received from each communicant recalled. Finally, they were asked to rate their confidence, on a scale from 1-7, about the information provided. At the end of each interview, informants were given the opportunity to send the experimenters a message containing any observations or suggestions they wished to make. Twenty-seven windows were established ranging from "one day, two days ago" to "one month, two months ago." Windows were selected for informants in random order. The remaining 10 windows we call "last on"; for these windows people were asked to recall their communications during the last time they were on EIES. This ranged from several weeks to several minutes in lag, and from several minutes to several hours in width.

Two questionnaires were also administered. The first interview collected data on all our informants' age, sex, self-reported EIES use, and self-reported estimates of memory ("how well, on a scale from 1-7, do you remember birthdays?"). The second interview was taken by the 22 informants who completed all 27 of the basic window interviews. It again asked for information on EIES use, and also asked informants to report the 20 people with whom they believed they communicated most. For each of those 20, informants were asked to rate (on a scale of 1-7) the importance of the communication, how satisfying it was, how desirable communication was with that person, and how interesting it was.

The data produced by this experiment are known as the EIES (pronounced "eyes") data; they are quite rich, and quite vast, offering many possibilites for measuring respondents' accuracy. (We have concocted 48 different measures of accuracy, most of which have been used previously in our series of papers.) A full report of the findings of this experiment are contained in TR #BK-120-80, which is still under publication review (Bernard, Killworth, and Sailer, 1980a).

c) Summary of Findings

In addition to the findings already cited, the comparisons between our informants' predictions of their behavior with their actual behavior (who they talked to on their TTYs) showed that 66% of all predictions made were erroneous. Furthermore, there was no way to predict which guesses were erroneous; there was no systematic effect on the accuracy of a respondent by any of the parameters we examined. These parameters included gender, amount of use of the TTY, number of communicants, length of time since acquiring a TTY, and so on. This left the unpleasant possibility that error in reporting behavior is produced by psychological or sociological factors which we will have to uncover before we can know the accuracy of any self-reported behavioral data.

Referees and other critics of our early work were very helpful, and quickly pointed out many apparent defects in our data. Among these were:

1. TTY communication, while natural to the deaf community, is not (on the face of it) a plausible proxy for other, more prevalent communication modes, including face-to face voice contact.[3]

2. TTY communication is essentially dyadic, whereas communication among people often takes place in groups. Does this affect accuracy?

3. The deaf community might have been giving cognitive data based on "typical communication" (e.g., an "average month") rather than on the actual three weeks under consideration; this would effectively magnify the observed error.

4. Different individuals might have been giving data based on various criteria (e.g. amount of communication, frequency of communication, importance of communication, etc.). This would produce gross inaccuracy when treated similarly.

5. Ranking individuals in a list may not be the most accurate way to collect data.

Perhaps asking for scaled data (i.e., "on a 1 to 5 basis, who do you ________?") would have revealed much less error.

6. Most of the communication in the deaf data took place outside the group. Perhaps people in a more fully closed group would be more accurate.

7. The data we collected were essentially precognitive. Perhaps postcognitive data would be more accurate. In other words, data about past events might be more accurate than data about future events.

As may be seen from the description of the data in subsection b), we have addressed each of these problems in at least one subsequent experiment. It is obvious, of course, that many replications of our experiment are required. Still, our work has produced monotonously similar findings. Intercomparisons among the various data sets yield the following results:

1. Postcognitive data are (mainly) more accurate than precognitive, but not significantly so.

2. With the curious exception of one's first ranked informant, there is not systematic variation in accuracy between asking for a "typical week's" data and "last week's" data.

3. There is no systematic variation in accuracy between data sets.

4. There do not seem to be any obvious personal, or socioeconomic data which have any bearing on accuracy.

5. Keeping (or using) communication logs does not improve accuracy significantly.

6. Asking people if they believe themselves to be accurate produces unreliable results.

7. There is (somewhat equivocal) evidence to suggest that informants judge on frequency rather than amount of communication.

8. Affective questions (e.g., "importance") are not systematically less accurate than effective questions which ask people to recall their behavior without regard to affective content.

9. Using the $\hat{A}$ accuracy score introduced in our first paper on this subject (see Killworth and Bernard, 1976), on average, over all data sets, people can recall or predict less than half their communication (measured on amount of frequency).

10. Even with a leeway of ±3, only the rank of the most-communicated-with person is reliably reported more than 50% of the time. The rank of the 2nd, 3rd..., 6th most-communicated, even with a ±3 leeway, cannot be relied upon half the time.

11. There is no evidence that any but a tiny percentage of communication can be accounted for by an informant's first "few" ranks (3, 5, 7, or whatever), or top "few" scales with any reliability whatsoever. Including more ranks or scales only makes matters worse.

12. Slightly obtrusive observation, such as occurs in behavioral sampling (the Tech data and Frat data, for example) has no noticeable effect on informant accuracy.

13. There is no obvious reason to prefer either ranked or scaled data on any measure of accuracy we have considered. Therefore, we recommend the use of scales on the grounds of convenience.

14. Telling people in a group that we expect them to get more accurate in repeated experiments over time produces no significant improvement in accuracy of reporting communication.

15. Attempts to predict communication from cognition (what one hopes one is doing by measuring cognition or recall) is not helped by including affect. In other words, how much i talks to j, as predicted by how much i thinks to talked to j, is not better predicted if one substitutes or includes knowledge of how much i says he likes j.

16. Although lag and width of the time window account for some of the variation in the accuracy of informants (small lags and width tend to be more accurate than large ones), the amount of variance accounted for is typically about 10 percent.

17. One positive finding emerged from our data: although people do not know with whom they communicate, people en masse seem to "know" certain broad facts about the communication pattern of a group. This may result from random errors in recall canceling each other out. But we don't know.

d. Structural Analyses

All our findings lead to one major conclusion: people do not know, with any acceptable accuracy, to whom they talk over any given period of time. Furthermore, the inaccuracy can not be accounted for by any of the usual characteristics of people or groups.

This leads to two interpretations. One is that there are two distinct networks, at least in communicative structures: cognitive and behavioral. Essentially, who people think they talk to and who people really talk are different networks, and should be treated as such. This may be true, but is hardly helpful if one is trying to study group structure: what one's instrument measures must have an existence -- or at least a correlate -- outside the bounds of the instrument itself, or else the instrument is useless. Of course, what we call cognitive data are statements by people about what they do. The correlate of these data may be simply what they think they do, with no correspondence assumed between an informant's thoughts about his or her behavior (say, communication) and his or her behavior. But then, what structure are we uncovering when we subject such data to analysis? If a group of 10 persons were all asleep and each person were dreaming of talking to at least one person in the group, then is there a group structure to be uncovered?

The other conclusion is that although the signal-to-noise ratio is extremely poor at the dyadic level, it may be somewhat better if one considers higher order structural elements.

Triadic level analysis

One step above the dyadic level of structural analysis is the triadic level of interaction. Holland and Leinhardt (1975) have provided the methodology for the examination of structure at the triadic level. Essentially a binary sociomatrix X_{ij} (where $X_{ij} = 1$ if i communicates with j, and = 0 otherwise) is scanned, and a triad census computed. This is a count of how many times each of the 16 possible triads occurs within the data (definitions of the triad types will be found in Holland and Leinhardt, 1975). The triads are distinguished by counts of the number of mutual, asymmetric and null dyads within them, together with other directional information when this is insufficient. An investigator, armed with the triad census, can then enquire whether some proposed structure (e.g., transitivity) occurs more often than chance in a set of data.

There are a great many possible structural building blocks which one might choose to examine.[4] We examined ten different potential structures: some familiar, like transivity and positive balance, and some created specially for the analysis, in order to make the point that many different kinds of structure do occur in data.

The following conclusions were drawn from our analysis of the triadic level of structure in behavioral and cognitive data:

1) There is an amazing amount of structure in both behavioral and cognitive data. There is so much structure, and the findings are so consistent (even for algebraic structures which we concocted just for the analysis) that one wonders what are the properties of triadic level structures which do not occur significantly often?

2) In the main, structure derived from ranked cognition data is very similar to structure derived from behavioral data treated as ranks. Similarly, structure derived from scaled cognition data is very similar to structure derived from behavioral data treated as scales. However, structures produced by ranking and structures produced by scaling are quite different. The methodological implications of this are obvious: how one treats data directly affects the qualitative and quantiative conclusions which may be drawn from it.

3) More than one set of structural tendencies can be drawn from the same set of behavioral data, depending on how it is treated numerically.

4) Since both behavioral and cognitive data showed similarly high counts of various (say, transitive) triads, this appeared to be an improvement in accuracy over comparisons by dyads in the data. However, this apparent increase in accuracy as the level of structure went up disappeared on close examination. In fact, when compared triad by triad, things got much worse. On average, any non-all-null behavioral triad is reported incorrectly 76% of the time. Thus, no reliance can be placed on the reporting of triads.

The Clique-level of Analysis

A step above the triadic level, we believe, is the clique-broker-link level of analysis; i.e., the level at which we assumed most of us consciously perceive group structure.

Cliques are typically obtained by applying some algorithms to cognitive or recall data. The assumption is that the cliques found in the cognitive data are those which would be found if one had corresponding behavioral data. It is quite possible that i states that he talked to j and k, when in fact he talked to l and m. This would produce great inaccuracy on both the dyadic and triadic levels of structure. But if i, j, k, l, and m form a clique, then i's report is a reflection of his interaction with that clique, though not its members. Thus a good clique-finding algorithm would be one which puts i, j, k, l, and m into a clique when applied either to cognitive or behavioral data. In other words, a good clique-finding device should reduce the noise which shows up as informant inaccuracy at the dyadic or triadic levels.

Since we possess matched pairs of behavioral and cognitive data on who talks to whom in a variety of groups, we were able to do a comparison on clique-finders. We chose three esnetially different and popular approaches: 1) factor analysis (Macrae, 1960); 2) an iterative correlational block modeling technique (CONCOR, see Breiger, Boorman, and Arabie, 1976); and 3) a graph-theoretic approach based on overlap of maximally complete subgraphs (COMPLT, see Alba, 1973).

There are many problems associated with comparing results from different clique-finders. First, all three of the algorithms which we chose (precisely because of their dissimilar approaches) differ in their data requirements. Most sociometric data are binary, while our data (collected by rankings and scalings) are not. Second, suppose that an algorithm is used on a matched set of behavioral and cognitive data; this produces two sets of cliques. How can we measure how similar such cliques are? Third, assuming that we have an adequate clique dissimilarity measure, how do we measure the difference between two sets of cliques (i.e., structure)?

A detailed account of how we treated our data, and the rationale for our dissimilarity measures for cliques and sets of cliques may be found in Bernard, Killworth, and Sailer

(1980b).

The three algorithms (COMPLT, CONCOR, FACTOR) were applied to four pairs of data (Office, Tech, Hams, and Frat). This produced twleve sets of comparisons betweeen behavioral cliques and cognitive cliques.

Given our definition of clique dissimilarity (see Bernard, Killworth, and Sailer, 1980b) the <u>best</u> dissimilarity (D) in the entire set is 0.50 (for COMPLT on Hams). The reader will immediately appreciate what this means:

> <u>For three major clique finders, run on four different sets of data, there is never more than a 50% concordance between the clique structure produced by people's recall of their interaction, and that produced by their interaction.</u>

Second, the different algorithms produce widely varying answers on the same set of data. The average "best" D, over all four data sets, is 0.89 (for comparison, the mean D over all comparisons is 1.6). The average D for COMPLT, CONCOR, and FACTOR were 2.18, 1.15, 1.48, respectively. (The variation between data sets is sufficiently large that no algorithm is significantly better than any other, on a one-way analysis of variance.).

Roughly speaking, then, the clique structure determined from a set of cognitive data differs 160% from the behavioral clique structure it is supposed to represent. For example, for any algorithm, the behavioral clique (1-2-3-4-5-6) is typically represented by the cognitive clique (1-7-8-9-10); this is, of course, the cognitive clique that <u>best</u> represents the behavioral clique.

We expected, at the outset, that Ds of 0.2 or so might occur; indeed, even 40% inaccuracy would be better than that seen at dyadic and triadic levels. After all, representing (1-2-3-4-5) by (1-2-3-4-6) was not, we felt, too bad a misrepresentation. A useful by-product of our clique level analysis, we had hoped, would be to find the alogrithm which most nearly fitted these reasonable demands on accuracy. But none did.

e) Discussion: where do we go from here?

It seems obvious to us that accuracy of clique representation could be improved by tinkering with default paramters, choosing individual cutoffs for binary data production, and so on. But how can a researcher know <u>a priori</u> how to do this? We are now convinced that cognitive data about communication can not be used as proxy for the equivalent behavioral data, at least at the dyadic, triadic, and clique levels of analysis. This leaves us with a problem, however, which must be resolved. Over the years, researchers have used their favorite clique-finding devies in order to provide managers with descriptions of the structures over which they (the managers) preside. Sociometry in the classroom is used in order to help teachers make decisions about groupings of children. Sociometry has been used in industry and in government to assess information flow in evaluations of productivity. Sociometric (or network) analyses have been used as the basis for the reorganization of task production units, and even for hiring and firing people.

We have used our own algorithm (called "catij," see Bernard and Killworth, 1973; Killworth and Bernard, 1974) in applied settings, and we have always found teachers, managers, and bureaucrats enthusiastic with the results. We ask people to rank order their communication with others; we produce a map of cliques and brokers between cliques; and we present the maps to members (usually managers) of the group. In one case, a colleague used catij to describe the structure of a tiny, isolated village in the mountains of Greece.

In all cases, the persons with whom we shared the maps offered spontaneous interpretations for the particular groupings, isolates, brokers, links, and so on. In other words, the maps <u>made sense</u> to our clients or village informants, even though (as we have shown) these maps could not have been even a close approximation of the actual dyadic, triadic, or clique structure of communication flow. Our colleagues report that, using their own favorite algorithm, <u>their</u> clients and informants are similarly enthusiastic with the

results. Their clients, too, respond immediately and spontaneously, putting the flesh of human explanation on the bare bones of the sociometric-cum-network maps placed before them. To make matters worse, as we have shown, different clique-finders produce very different results from one another. How can all this be reconciled?

We suspect that the answers may lie in discovering the regularities of a fourth level of structural analysis, the "folk" level.

When people say "members of clan A always marry members of clan B," they are engaging in folk structural analysis. When people in Ann Arbor say "there is a town-gown split here; the merchants and the university people simply don't know one another," this is a folk structural analysis. When academics say "graduates from school A are hired by school B, but not the other way round," they are making a folk structural anslysis. When the Purum of Burma explained the rules for cross-cousin marriage to Professor Edmund Leach, they were making a folk structural analysis. Everyone is familiar with the discrepancies between ideal, normative behavior (every man should marry his mother's brother's daughter) and reality (what does one do if one's mother has no brother?). People everywhere rationalize these differences, and create new rules for dealing with the problems created by old rules. Our next step, then, must be to conduct a series of investigations to see whether people can predict, as well as rationalize, ex post facto, the general form of the maps produced by clique-finders.

This is important if we are to construct a theory of information diffusion. Any such theory must be able to predict how information flows through the system, how quickly it will go from point A to point B, and how likely it is to be trapped in pockets and loops. This, it seems to us, is the goal of diffusion research. In order to address this goal, we have taken two approaches. The one described here, is an attempt to learn how to measure communication flow accurately. The other, described in Section III of this report, is an attempt to understand the decision making process by which information is retained or transferred along any of the multiple lines each of us has in our network.

For the future, we feel that a program of research is needed which will test the accuracy of many behavioral recall instruments. This must be done in may cultures, as well as in Western societies. We also need better measurements of communication per se. This means that we shall have to treat naturally occurring situations as experiments; and, above all, we must devise procedures for automated data gathering. (A crude, first approximation is the EIES experiment.) We will have to concentrate on the two ends of the methodological spectrum: the essentially unverifiable ethnographic method may allow us to understand how people deal with and organize the overwhelming data of communications reality; the automated experimental technique may allow us to describe that reality. From our work thus far, we are convinced that the more convenient, intermediate methods (questionnaires, card sorts, and other forms of behavior recall prods), produce too much error to be a proxy for either the folk level or the behavioral level of reality. Furthermore, the error is so great, that statistical and numerical techniques for washing data collected by recall instruments, can not solve the problem.

III. Small-worlds, reverse small-worlds, and their role in social structure

a) Introduction

The diffusion of information, innovations, a contagious disease, or whatever, through some population has been thoroughly studied by social scientists for many years, dating back at least to Tarde (1903).

The classic diffusion study of Coleman et al. (1966) provided the impetus for diffusion researchers to ask sociometric questions of the members of the social system. Leaving aside the obvious problems of whether or not an individual knows the answers to the sociometric questions to any useful degree, have sufficient sociometric data been obtained to be useful in interpreting diffusion studies? Have the right types of data been obtained? How does one acquire the right type, whatever that might be? And so on.

There seems to be little doubt that the more we understand social structure (here defined as the patterns of who knows whom), the more likely we will be able both to predict diffusion (of, say, ideas) and also to interpret the diffusion data themselves.

So, how should one acquire data about real-world social structure? At first glance, all we need to do is ask each member of the structure for a list of all the people he or she knows. With unlimited patience, a huge computer, a lot of luck, and assuming all informants managed to remember the thousands of people they knew, this procedure would suffice; in real life, of course, it would be disastrous.

Clearly we need to find out both *less* and *more* information than this. We need *less* information because many of one's acquaintances serve no useful purpose for us in our lives; we only need to determine the acquaintances who are, in some sense, useful. (Exactly what is meant by useful is rather difficult to define.) We need *more* information because, of an individual's useful acquaintances, we need to know how and why that individual knows them. For example, a farmer in Iowa may have a best friend in Kuala Lumpur. By no stretch of the imagination can, say, a contagious disease spread directly between them; but a snippet of information can. So just knowing links in a network of acquaintances tells us little unless we also know something about those links.

Traditional sociometric tools are, as we know from our work on informant accuracy, inefficient at gathering this kind of information. We approached this problem by considering the small-world method, due originally to Milgram (1967), and how we might improve on the method in order to produce the kind and amount of information required for a theory of social (i.e., communications) structure. The resultant methods, invented under this contract, are known as the reverse small-world, and the informant-defined reverse small-world method.

b) The small-world method: the experiment

Although Milgram's now classic 1967 experiment began the accepted chain of small-world (henceforth SW) papers, the origins of the problem it was designed to solve lie in 1958. In that year, Pool & Kochen wrote a paper which circulated rapidly through the academic underground, finally reaching publication in 1978. They - and Milgram - were interested in the answer to a deceptively simple question: "starting with any two people in the world, what is the probability that they know each other?" The probability, about 5×10^{-5}, wasn't very enlightening, so the problem was expanded: "given any two people in the world, person X and person Z, how many intermediate acquaintance links are needed for X and Z to be connected?" Pool & Kochen (1978) had already estimated that 50% of such pairs could be connected by two intermediate links (assuming, of course, that X and Z were aware of these connections, itself an unlikely event).

This problem proved tractable by one of the most elegant (and cheap, the total cost being $680) of all social science experiments. Milgram created a pool of *starters* (henceforth Ss). These were individuals in various parts of the U.S. who were prepared to help with the experiment. There were two groups, functioning independently: 145 persons in Kansas and 160 persons in Nebraska. Each S in each group was given a folder containing some background information (name, address, occupation, marital status, etc.) about a *target* (henceforth T) person. The T for the Kansas group was the wife of a divinity studnet in Cambridge, Massachusetts; the T for the Nebraska group was a stockbroker in Boston.

The Ss were given the task of getting the folder to the appropriate T through a chain of acquaintances, as rapdily as possible. In other words, each S chooses the person that S thinks is most likely to know T (or most likely to know someone who is most likely to know T, etc.) and gives or send the folder to that intermediary. The intermediary then effectively becomes a new S and the chain continues, until one intermediary either actually knows T or, for some reason, drops out of the experiment.

Milgram (1967, 1969) wanted to know such things as: how many steps, on average, it took to get from any S to the T?; and were there qualities of the Ss or Ts which affected

the number of steps involved? However, we might also think of the experiment as an attempt to discover how many people know T in a "useful" sense. After all, those people in the chains who actually passed the folder to T from a well-defined group: they are a (subset of the) class of people who know T. Would this be a large or small group? The questions were intriguing, and the freshness of the method generated a great deal of interest among structural theorists.

c) The small-world method: results

Of all the SW chains initiated by Milgram, only 44 were completed (this appalling attrition rate - about 25% per step of the chain - is typical of "real world" experiments). Remarkably, the average chain length from S to T was 6.2 steps, with a mode of 7.

Knowing a mean path length tells us little about social structure, of course. Travers & Milgram (1969) performed the first expansion of Milgram's original experiment by using the same T for two groups of Ss: one in the same city (Boston) as T, and one half a continent away. The "local" chains were significantly shorter (5.4 vs. 6.7). Obviously, some form of social distance, with a geographical component, is at work here.

Perhaps the most useful of their findings for social structure - and, indeed, of the papers which followed, which are reviewed elsewhere (Bernard and Killworth, 1979) - was a clustering effect observed as chains neared T. Forty-eight percent of all chains reaching T in their study came in through just three penultimate links. So incoming networks are highly structured. Presumably outgoing networks are, too (i.e., if T was to serve as a starter to many new targets, he might choose some intermediaries very often and others hardly at all).

Although the SW technique had, by the late '70s, been performed in businesses, multinational dormitories, and the like, one gets the impression that the information obtained is not in quite the best form to use in a theory of social structure. The (repeatable) facts that SW experiments produce are the *results* or output of the social structure, and it is not easy to see how to plug these back into a theory. In fact, it is fair to say that models of the SW experiment have yielded more information germane to social structure than the experiments themselves. We generated a model of Milgram's experiments (Killworth and Bernard, 1979), and created a flow chart which we felt represented the thought processes undergone by a participant in a SW experiment.

Surprisingly, almost all the predictions of the model agreed with observations. For example, we computed the mean complete path lengths to T from three categories of starter: far (not in a circle of 7×10^6 population), far but occupation-connected, and within 7×10^6 population with S at the center of the circle. The path lengths were 6.5, 5.8, and 5.0 respectively; Travers & Milgram (1969) found 6.7, 6.4, and 5.4 respectively in their experiments.

The SW technique is thus generating ideas for models of social structure. But can the method be adapted to yield richer, more directly applicable data? We felt that it could, and the result was the reverse small-world experiment (RSW).

d) The reverse small-world method: the experiment

No matter how many Ss one uses, one obtains, per target, three pieces of information: a) how many people comprise his incoming network, assuming an awful lot of starters were used; b) the mean path length to that T, and hopefully a fit to various SEC indicators of Ss and Ts on this; and c) scattered snippets about intermediaries in the chains. To get all this requires vast resources, due to the attrition rate, and therefore a concomitant increase in complexity and cost.

We (Killworth and Bernard, 1978a) attempted to avoid these problems by eliminating the SW task entirely. Instead of many Ss and one T, we decided on fewer Ss but many Ts, with no passing of folders involved. We created a long list of mythical targets (1267 in

total). First and last names were paired from different telephone directories; 168 names were suitably ethnic in origin (e.g., Wong Fuk Lam). An address was provided for each T, roughly reflecting the U.S. population; 100 were foreign (i.e., non-U.S.); 100 were local (i.e., in the two neighboring states to West Virginia, where the experiment took place, and in West Virginia itself). Half were male, half female. Each T had an occupation, spanning the Duncan (1961) scale. The list was then shuffled, printed up as an instrument, and presented to starters.

Each of the 58 starters, who were paid for their lengthy (8 hours apiece) participation, considered each T on the list. Armed with the knowledge of T's name, location, occupation, ethnicity (Blacks were indicated) and sex (unless the name was Oriental), each S made his or her choice for an intermediary in a SW chain to that T. They provided the choice's name, and relationship to S (friend, acquaintance, or one of 21 types of relative), and checked one of four possible reasons for making that choice: location, occupation, ethnicity, or other (the latter being left unspecified).

This provided two sets of interrelated data. We had a list of targets about which we knew occupation, sex, ethnicity, whether they lived in a large or small town, and location by state (or country). We also had, per starter, a list of choices, one per target, about whom we knew sex, relationship to starter, and a reason for choice. Additional starter information (e.g., their sex, age, income, religion, etc.) was also obtained.

e) The reverse small-world method: results

Of immediate interest was the size of an individual's network, i.e., how many different choices were generated by the list. The average number was 210 (ranging from 43 to 1131), with a highly skew distribution (all but two Ss making less than 400 choices). Monitoring the mean number of choices generated for the first n targets on the list, as n increases, we found that at the end of 1267 targets, the number of choices was still increasing. (The shape of the curve refused to fit any plausible model assumption, unfortunately.) We estimate that over 2000 U.S. targets and 500 foreign targets would be necessary to exhaust an individual's network.

Some choices were far more "popular" than others. On average, only 35 choices were required to account for half of all the targets (and only 3 for 10% of the targets). Similarly, choices were used more often for one reason than another: 45% were chosen most often for location reasons, 47% for occupation, and only 7% of choices were mainly based on ethnicity or other reasons.

Eighty-two percent of the time a friend or acquaintance was used for a choice (the terminology differs according to sex of starter, with males preferring acquaintances and females friends). For any given target, indeed, the type of choice used most often was never a family member.

Characteristics of Ss and Ts enabled many strong predictions to be made. For example, the most likely sex of the choice, for any given target, can be predicted accurately 82% of the time. The sex is male, unless both starter and target are female, or if the target has a low-status occupation. Similarly, the most popular reason for a choice, for a given target (always location or occupation) can be predicted accurately 81% of the time. Essentially, location is preferred as a reason except for targets with a high-status occupation or in faraway small towns. This preference agrees very well with the experimental results of Travers & Milgram (1969) for their stockbroker target (occupation level 85 out of 100).

We were also able to quantify such common phrases as "one's man in Idaho." Virtually never was a single choice used for every target in a single U.S. state. However, the choice accounting for *most* of the Ts in any state (when location was the reason) accounted for 69% of those Ts. Defining a choice to "handle" a state (i.e., to be "a man in . . .") if that choice accounted for two-thirds or more of the Ts in that state, we found that almost half the states in the U.S. were handled by a (usually different) single person.

This suggests that most Ss have some kind of cognitive map of the entire U.S., which they tap when approached by social scientists requesting information (do they use this map in normal life?, or is it an artifact of our experimentation?). What makes some targets seems like others to an S, who proceeds to use the same choice for them both? As an experiment, we considered the first 100 Ts. We argued that if a starter used the same choice for two Ts, he or she perceived that pair as similar, and that the more Ss who did the same, the more similar that pair of Ts were. We performed a multi-dimensional scaling on such a matrix of similarities, placing the 100 Ts into a two-dimensional space to that similar Ts were close, and dissimilar Ts were far apart.

The resulting map was rotated to resemble a genuine map of the U.S. The resemblance is certainly very good, with the South, New England, and Far West states basically placed correctly. California and Texas were misplaced, and high-status targets migrate toward the edge of the diagram. Clearly there is some kind of - not necessarily geographical - map in informants' heads.

One of the problems with RSW was the size of its data (at least 73,000 informant-target pairs alone). There is a tendency to assume that because the fits to the data accounted for so much variance, and because we had so much data, that the *quantitative* results should be applicable elsewhere, but the INDEX experiment described later failed to fit these results. However, there seems little doubt that all the qualitative results should hold for further data, with minor parameter adjustments, etc.

We felt, then, that RSW had provided a great deal of information about some aspects of social structure, and an embarrassingly high number of significant straight lines in the data. But yet there were at least two shortcomings. First, we had gathered very little information about the choices. We knew their names (and therefore, usually, their gender) and whether they were relatives or friends of the informant. But if all the choices made by an S are to be something meaningful, then there must, we hope, be a pattern amongst those choices; something that announces this group of people to be connected with S. The sheer labor of investigating 210 x 58 = 12,180 choices after the experiment does not permit such analysis. So we know - at *least statistically* - why S chooses certain types of people for certain types of T; but not why S knows them in the first place.

Second, the RSW instrument was closed-ended. It provided only a few pieces of information about each target, because SW experiments do. In turn, SW experiments provide such information because accepted sociological theory tells us such information is important. For the same reasons, informants were only allowed to check certain reasons for their choice, whereas the actual reason might be very complex.

In fact, informants' comments about the RSW experiment revealed two interesting details. They occasionally asked about other target information which was not provided; and frequently (when we checked) choices were made on the basis of location who had never lived anywhere near T. But informants claimed their choices to be *associated* with T's location because, for example, the choice's children might have gone to college in the same town as T lived. (This led to the model discussed earlier.)

If this is the case, then might it not work in reverse? If an S were told where T's children had gone to school, might that not be of use to S in making a choice? Or knowing T's hobbies might be useful, or . . . the list was endless.

By this time, it was obvious that the only persons who knew what information about T was required would be the starters themselves. This led us (Bernard, Killworth and McCarty, 1980) to perform an "informant-defined experiment," or INDEX. The idea is to study social structures experimentally, but to allow the subjects of the study to define the information which is collected.

f) The informant-defined small-world method: the experiment

A great deal of pretesting revealed that the list of targets had to be kept short,

both to maintain informants' interest and to prevent the data from getting out of hand. We settled on 50 targets, all mythical. Each target was assigned a name, gender, occupation, location and a racial identity as before. New information was then added for each target: an age, a religion, an education level, and marital status. After pretesting, a maximum of five hobbies and five organiztions were added, together with details of number of children, etc.

The reverse small-world procedure was explained to each of 50 informants. We explained that we had complete life histories of 50 people from around the U.S., but with names and characteristics shuffled to protect anonymity. Targets were presented in a random order, to minimize learning effects. Informants were given no information about any target. However, they were instructed to ask any questions they liked about each T; the questions were all answered.

Of course, it frequently occurred that a question was asked without an answer in the target's dossier. Either the informant was told the information was not available, or else (more frequently) it was made up on the spot, and later added to the dossier. There were, predictably, many problems with this procedure, and these are discussed in detail in Bernard, Killworth, and McCarty, 1980.

Each question ever asked was assigned a unique number, with no connotation as to order. For example, question 3 refers to target's occupation, and question 14 to target's location. For each target, then, the code number of each question asked was recorded in sequence. When informants had asked enough questions, they stated their choice. Then they provided a "few sentences" which explained why they had selected that choice (i.e., "because he's a real estate agent," or "because his girl friend's father is a pharmacist"). Next, informants ranked the questions they asked by the degree to which the answer had helped them make their choice. They were required to select a first-ranked question, and could rank up to four more. All other questions were graded by the informant as "helpful" or "not helpful." The relationship of each choice to the informant was recorded.

Finally, after completing the test, each informant answered a questionnaire. This consisted of basic sociometric data, and a personal response to any question ever asked by the informant about any target.

Most of this information thus presented was straightforward to code (with reservations about location, for which we provided five distinct definitions). The problem lay with the "few sentences." Four concepts were introduced, the "direct hit," the "associated hit," the "via," and the "intervening choice." If an explanation revealed that a characteristic of a choice matched exactly to a characteristic of the relevant target, this was a direct hit. For example, if a target lives in Los Angeles and the choice lives in San Francisco, then if, and only if, the informant said he selected the choice on the basis of location, this counts as an "associated hit." Associated hits can occur for a wide variety of reasons. If an informant says he chose a pharmacist in order to get to a physician because "they are both in the medical field," then this is an associated hit. Similarly, a farmer and a tractor salesman may be associated by occupation; a student choice may be associated with a college administrator; a choice who plays a jazz trumpet as a hobby may be associated with a target who collects jazz records, and so on. The concept of "associated location" and "associated occupation" has been introduced earlier. Our experience in this experiment has broadened the concept to include associations such as hobbies, organizations, religions, etc.

In fact, our experience with these data has shown that simple associations are not enough to describe all the relationships which informants claim exist between their choices and the targets. This led to the "associated via" and "intervening choice" categories. Consider the case of a choice who is a coal miner linked, by an informant, to a target who lives in Kentucky. The coal miner choice may, in fact, live in Ohio. But if the informant says, "I chose him because he is a coal miner and he could contact people in Kentucky where there are lots of coal miners," then we believe this is best described

as "associated with target's location via choice's occupation." Some other examples include the following: "I chose her because she belongs to the Sierra Club and the target works for the Environmental Protection Agency," then this counts as "associated with target's occupation via choice's organizational affiliation." "I chose him because he does cross-country skiing and the target lives in Vermont" is coded as "associated with target's location via choice's hobby." "I chose him because he collects rocks and the target is a geology student" is coded as "associated with target's field of study via choice's hobby."

Finally, many of our informants were apparently thinking two steps into the small-world problem when they said such things as "I chose him because his girlfriend worked at Kroger's grocery and the target owns a grocery store." This counts as "associated with target's occupation via intervening choice's occupation." The *choice* was not associated with the target by any characteristics of his own; but his girlfriend (whom the informant may not have known well enough to name as his choice) is associated with the target's occupation. For simplicity, we code the fact that the girlfriend is an intermediary choice, and that she is somehow associated with the target's occupation. Another example is the following: "I chose her because her father used to be a professional pool hustler. He could contact the target who likes to play pool." This was coded as "associated with target's hobby via intervening choice's occupation."

g) The informant-defined reverse small-world method: results

As we had hoped, the two most frequently asked questions (out of 82 different questions created by informants) were indeed target's occupation and location (asked, respectively, on 92% and 90% of all occasions). Other questions were much less frequently asked: age of target (42%), sex (36%), marital status (24%), and hobbies (21%). Put another way, location and occupation together contributed 38% of all questions ever asked; age and sex, when added, contribute over 50%.

Furthermore, *the dominance of location-occupation* continues if one examines "most helpful," "at all helpful," or even "unhelpful" questions. This lack of dependence on whether questions are useful suggests that the same questions tend to be asked about all targets. However, the distribution of the "most useful" questions differs subtly from the others: location-occupation account for 64% of all "most useful" questions, with hobbies and organizations raising the total to 75%. We have thus concluded that the basic set of questions:

target's location
occupation
hobbies
organizations
age
sex
marital status

supply the basic information about any U.S. target (to an S who lives in the U.S., at any rate). Name of target does *not* feature on this list.

We were able to show fairly accurately how a string of questions is created by an S, following a flowchart. Even when a question like sex of target is asked first, informants find it necessary to ask location and occupation and then proceed on the basis of how useful the results of such questions were. The later stages of all such flowcharts are all very similar.

Thirty-five different probabilities (e.g., that location is the most useful; that marital status is not useful, etc.) can be described with more than 40% of variance accounted for (up to 71%, in fact) by linear combinations of target data. Here target characteristics control most of the questions which informants ask.

The choices made by Ss (i.e., family or friends) and their sex, again reflect the findings of RSW. Of the 50 targets, location was the most popular reason 23 times and occupation 25.

Surprisingly, the probability of a direct hit (as defined earlier) is 90%. Of course, location and occupation were the most likely to be direct hits (19%, 15% respectively). The chance of a location direct hit was fitted (75% of variance) by target characteristics: the nearer and more urban the target, the more likely a direct hit.

Similarly, associated hits have a 95% chance of occurring, with corresponding vias at 88%. Location-occupation accounted for over 60% of all associated hit. Intervening reasons (and vias) are distinctly less likely, at 10% probability. Thus in all cases, location-occupation retains its dominant role in question and choice selection.

We assumed that S selects a choice for a given T because, in some sense, S perceives the choice to be similar to T. Furthermore, given several similar choices, S chooses the most similar such choice. How can we model the complex cognitive processes yielding such a similarity? As a simplification, we assumed a choice and T to be perceived as similar if and when some facet of the choice (e.g., where the choice went to school) and some facet of the target (e.g., where one of T's children lives) are either connected, or, at best, identical.

Each such facet of a target's personal history we term a "tag." On average, targets developed 16 tags (we counted tags in each of location, occupation, hobbies, organizations, age, sex, and religion, the latter three categories having one apiece), with 5 given over to locations. We deduced choice tags from the question responses, since we knew nothing else about the choices. On average, choices have two tags (but one had 12).

This enables a test of a simple hypothesis. We can predict the most likely choice for a given T by comparing tags until we find maximal agreement. The procedure is biased, of course, by the backward way of discovering the choice tags, but this is allowed for statistically. We measured the accuracy of the model by "easy" and "difficult" scores. The easy score is unity whenever the actual choices are among the optimal choices, and zero otherwise. The difficult score is 1/(number of optimal choices) if the actual choice is among the optimal choices, and zero otherwise. In other words, the easy score counts how often the actual choice was correctly (but not necessarily uniquely) predicted; the difficult score counts how often we would be correct if we chose at random among optimal choices.

The model works well with an average easy score of 89%, and a difficult score of 60%. Both are significantly (better than 1% level) higher than expected by the biased way the data were calculated. However, no weighting of tags (either by direct or indirect hits, or by giving more weight to, say, location tags, or whatever) improved the accuracy. As defined here, all tags have an equal utility. We deliberately did not restrict the target's tags to what each informant knew of the target (i.e., we compared choice location tags with a target's places of travel whether or not the informant had asked about T's travel) as this would further - but artificially - increase accuracy.

h) Conclusions and future research

It may be that the kind of data we seek (i.e., for all members of a group: who does each member know, and why?) are far too unwieldy to elicit any firm laws about structure. After all, the motion of a liquid or gas is best understood at the bulk motion level, and not by considering the quantum dynamics of each atom in turn. Perhaps in small-world studies we are still (incorrectly?) looking at the atomic level of structure. If this is so, can we achieve the bulk motion level by simple averaging over people?

We also need to know more precisely what information about a target is needed to "define" that target to an informant. The list of questions we presented, after all,

relates specifically to the experiment we performed. But at least one could test this, in a small-world context, very easily. One creates three sets of SW experiments, all with the same target. One group of Ss is given the answers to *all* the different questions ever asked in the INDEX experiment; one group can request information just as in the INDEX experiments; and the last group (which could be subdivided) is given just T's location and occupation (or perhaps hobbies, etc.). Then one examines whether SW chains differ significantly in length; either way one learns something, whether they do differ or not.

So perhaps we can define the essentials of a target - at least for basically Western European informants. (There is an obvious need for cross-cultural comparison - provided we know what to compare. The concept of a "useful" choice probably differs between the U.S. and a Mediterranean culture, for example. How can we handle this, let alone account for it?)

But how can we define an informant as a unit in the structure? After all, something as useful as the tag concept still founders when one askes "*what* makes Ss have more, or different, tags than others?" It is simply not good enough to blame "personal history of informants" for this failure of basic SES variables to account for differences in tags between informants. And yet aggregating our informants (i.e., ducking the problem entirely!) may not be the answer. Over what group of informants should one aggregate? All Bostonians? All violin players? All the U.S.? Just because these subgroups make (occasional) sense to us doesn't mean they are correct, after all. But surely we don't need to factor analyze data from the whole world population to find how to aggregate?

Obviously what is desperately needed are testable, falsifiable *theories* of social structure. The falsifiable criterion is vital. Heider's balance theory still has its proponents despite its refusal to occur in data; so small-group research needs better theories. We assume tacitly that a theory, however unlikely or implausible - an awful lot of physical science is thoroughly implausible - can be modeled so that predictions can be made, and tested.

But what predictions should be made, and why? (Granovetter, 1979, raised the same awkward question.) We suspect that at this stage in our knowledge, or practically any subject, let alone social structure, we do not really know (a) what we would do with perfect, complete, noise-free data, and (b) how we should compare that data with theories.

This is certainly pessimistic. Now in some scientific areas (e.g., meteorology) we have some very practical ways of checking predictions: did it rain today, like the compueter said it would? But in the social sciences, except those based firmly in the public domain, we have been content for too long merely to *describe* the situation. Perhaps now the pendulum is beginning to swing back, and we shall try to *understand* and *predict* what is happening in a real social structure.

NOTES

[1]Some scholars argue that what people talk about is as important as how often they talk to each other, or for how long they talk. We take the position, using Occam's razor, that the content of conversation is not demonstrably important in understanding structural change in human relations, and that it is not measurable. We claim that amount and duration of interaction between persons is measurable (or ought to be), and until it can be shown that measurable quantities do not yield adequate data about social structure, there is not reason to cloud the field further with attempts to include meaning

[2]An aside is in order here. It has been suggested that giving an informant a list of names may well influence who he actually contacts (though not, presumably, those who contact him). This may be true. It is also likely that asking children in a classroom for their three favorite friends will influence their later behavior, but this is usually ignored in the literature. Any act of data gathering must induce a quantum jump in the system being observed, whether the system be a social network or a hydrogen atom. The difficulty arises because one can compute the expected magnitude of the jump for a hydrogen atom, but not for a social network. It is not obvious, in other words, how both to obtain data and to stop informants thinking about their choices afterwards. Leo Tolstoy, as a boy, believed that any wish would be answered if only, after making it, he could stand facing a wall and not think of a white bear.

[3]We are constantly amazed at this criticism, because it comes up so consistently in much social science literature. In physics, a finding about the behavior of waves in the Baltic Sea would never be faulted on the grounds that "the Baltic is not typical of seas." Such a criticism, in fact, would be absurd. (It may be the case that wave forms in one sea are different from those in another.) Is eating with a fork or chopsticks more "typical" of current human behavior? If chopsticks are more typical, then are forks abnormal?

[4]See Killworth and Bernard, 1979a, for a discussion of how we converted non-binary data into a sociomatrix. Scaled and ranked data, of course, must be treated differently. More importantly, however, there is more than one way to handle behavioral (or any valued) data in order to produce a sociomatrix. It turns out that different conversion techniques produce widely differing structural tendencies. How the data are treated, alas, determines the answers one gets from the analysis.

REFERENCES

Alba, R.
1973 A graph-theoretic definition of a sociometric clique. Journal of Mathematical Sociology 3:113-26.

Bernard, H. R., and P. D. Killworth
1973 On the social structure of an ocean-going research vessel and other important things. Social Science Research 2(2):145-84.
1977 Informant accuracy in social network data II. Human Communication Research 4(1):3-18.
1978 A review of the small world literature. Connections 2(1):15-24.
1979 A review of the small-world literature. Sociological Symposium no. 28, Fall, 87-100.

Bernard, H. R., P. D. Killworth, and C. McCarty
1980 The reverse small-world experiment II. Technical Report #BK-119-80,ONR Code 452.

Bernard, H. R., P. D. Killworth, and L. Sailer
1980a Informant accuracy in social network data IV. Social Networks 2(3):191-218.
1980b Informant accuracy in social network data V: an experimental attempt to predict actual communication from recall data. Technical Report #BK-120-80, ONR Code 452, contract #N00014-75-C-0441.

Breiger, R. L., S. A. Boorman, and P. Arabie
1975 An algorithm for clustering relational data with applications to social network analysis and comparison with multidimensional scaling. Journal of Mathematical Psychology 12:328-83.

Coleman, J. S. et al.
1966 Medical innovation: a diffusion study. New York:Bobbs-Merrill.

Coleman, J., E. Katz, and H. Menzel
1957 The diffusion of an innovation among physicians. Sociometry 20:253-70.

Davis, J. A.
1979 The Davis/Holland/Leinhardt studies: an overview. In Perspectives on social network research. P. W. Holland and S. Leinhardt, eds. New York:Academic Press.

Duncan, O. D.
1961 Socioeconomic index scores. In Occupations and social status. A. J. Reiss, Jr., ed. New York:Free Press.

Granovetter, M.
1979 The theory-gap in social network analysis. In Perspectives on social network research. P. W. Holland and S. Leinhardt, eds. New York:Academic Press.

Hiltz, S. R., and M. Turoff
1978 The network nation. Reading, Mass.:Addison Wessly.

Holland, P., and S. Leinhardt
1975 The statistical analysis of local structure in social networks. In Sociological methodology. D. R. Heise, ed. San Francisco:Josey-Bass.

Hunter, J. E., and R. L. Shotland
1974 Treating data collected by the "small world" method as a Markov process. Social Forces 52:331-32.

Killworth, P. D., and H. R. Bernard
1974 Catif: a new sociometric and its applications to a prison living unit. Human Organization 33(4):335-50.

1976 Informant accuracy in social network data. Human Organization 35(3):269-96.
1978a Help! A dynamic social network data base. Connections 2(1), March:10-11.
1978b The reverse small-world experiment. Social Networks 1:159-92.
1979a Informant accuracy in social network data III. Social Networks 2(1):19-46.
1979b A pseudomodel of the small-world problem. Social Forces 58:477-505.

Macrae, D., Jr.
1960 Direct-facotr analysis of sociometric data. Sociometry 23:360-71.

Milgram, S.
1967 The small-world problem. Psychology Today 1:61-67.
1969 Interdisciplinary thinking and the small-world problem. In Interdisciplinary relationships in the social sciences. M. Sherif and C. W. Sherif, eds. Chicago: Aldine.

de Sola Pool, I., and M. Kochen
1978 Contacts and influence. Social Networks 1:1-48.

Tarde, G.
1903 The laws of imitation. New York:Holt.

Travers, J., and S. Milgram
1969 An experimental study of the small-world problem. Sociometry 32:425-43.

DA
FILM